

OPEN

# Association of relative brain age with tobacco smoking, alcohol consumption, and genetic variants

Kaida Ning<sup>1,2</sup>, Lu Zhao<sup>1</sup>, Will Matloff<sup>1,3</sup>, Fengzhu Sun<sup>2</sup> & Arthur W. Toga<sup>1\*</sup>

Brain age is a metric that quantifies the degree of aging of a brain based on whole-brain anatomical characteristics. While associations between individual human brain regions and environmental or genetic factors have been investigated, how brain age is associated with those factors remains unclear. We investigated these associations using UK Biobank data. We first trained a statistical model for obtaining relative brain age (RBA), a metric describing a subject's brain age relative to peers, based on whole-brain anatomical measurements, from training set subjects ( $n = 5,193$ ). We then applied this model to evaluation set subjects ( $n = 12,115$ ) and tested the association of RBA with tobacco smoking, alcohol consumption, and genetic variants. We found that daily or almost daily consumption of tobacco and alcohol were both significantly associated with increased RBA ( $P < 0.001$ ). We also found SNPs significantly associated with RBA ( $p\text{-value} < 5E-8$ ). The SNP most significantly associated with RBA is located in MAPT gene. Our results suggest that both environmental and genetic factors are associated with structural brain aging.

The number of American aged 65 and over is projected to reach 80 million by year 2050<sup>1</sup>. The brain aging process, while associated with structural changes, declined cognitive function, and increased risk of dementia, differs between individuals<sup>2-4</sup>. Therefore, to understand the factors associated with brain aging becomes increasingly important.

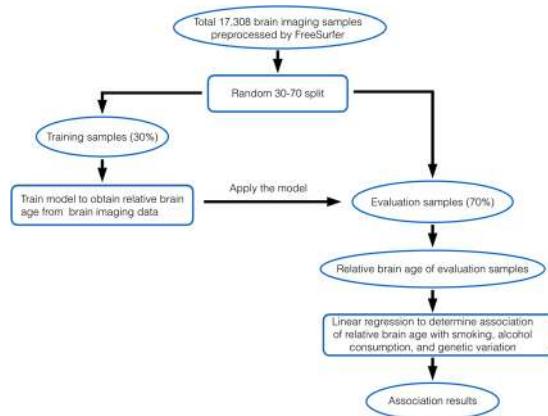
It is known that certain lifestyle habits are associated with accelerated atrophy in specific brain regions. Heavy smoking and heavy alcohol drinking are among the most studied adverse factors. Compared with non-smokers, smokers have significantly smaller grey matter volume and lower grey matter density in the frontal regions, the occipital lobe, and the temporal lobe. Further, smokers have a significantly greater rate of atrophy in regions that show morphological abnormalities in the early stages of Alzheimer's disease<sup>5-7</sup>. It has also been reported that patients with alcohol use disorder show decreased regional grey and white matter volumes in the medial-prefrontal and orbitofrontal cortices. The loss of brain gray and white matter volume accelerates with aging in chronic alcoholics<sup>8,9</sup>. On the other hand, studies have shown that nicotine, a compound contained in tobacco, may improve attention and other cognitive functions in human subjects<sup>10,11</sup>. It has also been reported that drinking wine may be beneficial to the cardiovascular system, which is related to brain health<sup>12,13</sup>. To date, it is still unclear how smoking and alcohol consumption is associated with brain structural aging, especially when the morphology of all the brain regions is considered.

Besides lifestyle habits, genetic factors are also thought to be involved in brain aging. A recent study analyzed brain imaging data and chronological age (CA) information from twins and suggested that the brain aging process was heritable<sup>14</sup>. However, the extent to which individual genetic variants are associated with brain aging hasn't been well studied, except for some conflicting results regarding the association between genetic variation in APOE, a gene associated with Alzheimer's disease, and brain aging<sup>15-17</sup>. Therefore, we also investigated if genetic factors are associated with brain aging in addition to smoking and alcohol consumption. Further, genetic and environmental factors might interact in affecting traits, risk of complex diseases, or lifespan<sup>18-20</sup>. For example, hypertension interacts with APOE  $\epsilon 4$  risk allele in affecting cognitive function<sup>21</sup>; the diabetes drug metformin extends lifespan of the rat model with hypertension, yet doesn't extend lifespan of healthy rats<sup>22</sup>. Therefore, it is important to explore if there is genetic and environmental factor interaction in association with brain aging.

<sup>1</sup>USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of University of Southern California, Los Angeles, California, 90033, USA. <sup>2</sup>Molecular and Computational Biology Program, University of Southern California, Los Angeles, CA, 90089, USA. <sup>3</sup>Neuroscience Graduate Program, University of Southern California, Los Angeles, CA, 90089, USA. \*email: [Toga@loni.usc.edu](mailto:Toga@loni.usc.edu)

	Number of subjects	Male (%)   Female (%)	Age (mean [SD], min–max)
Training data (for model training)	5,193	2,466 (47%)   2,727 (53%)	63.3 [7.4], 46.2–80.7
Evaluation data (for association analyses)	12,115	5,753 (47%)   6,362 (53%)	63.3 [7.4], 45.2–80.3

**Table 1.** Demographic information for subjects included in the training and the evaluation data sets.



**Figure 1.** Procedure for studying the association of relative brain age with smoking, alcohol consumption, and genetic variation.

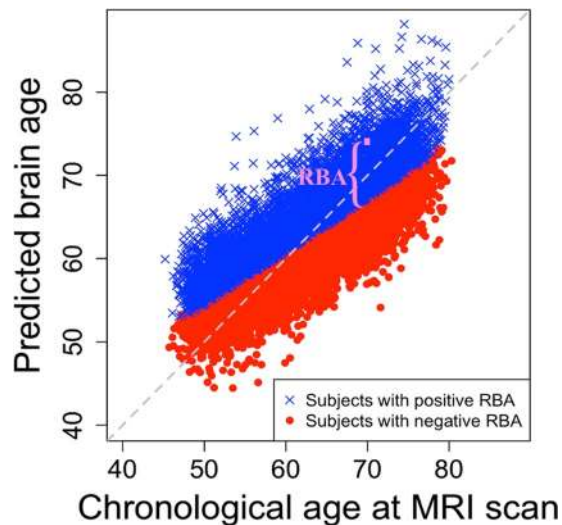
Recently, researchers have successfully used machine-learning methods to derive a biomarker that is commonly referred to as predicted brain age (PBA) or brain age based on brain imaging data. PBA reflects the degree of aging of the brain based on its anatomical characteristics, as computed based on brain morphology measurements across the entire brain. PBA has been derived and used in several studies, where the mean absolute error between PBA and CA was less than 5 years in adults<sup>14,23,24</sup>. Further, it has been shown that advanced brain age is associated with Alzheimer's disease, objective cognitive impairment, and schizophrenia, etc.<sup>23–27</sup>. Before our research, many papers used the difference between PBA and CA (i.e., PBA - CA) for capturing deviation of person's brain structural aging from norm<sup>15,17</sup>. However, due to regression dilution, this metric is correlated with CA and may not be optimal<sup>28,29</sup>. Therefore, we further developed a metric called relative brain age (RBA), which is independent of CA and indicates if a subject's brain has experienced accelerated or decelerated aging compared to peers. While our manuscript was under review, Smith *et al.* independently reported a method for improving brain age delta estimation<sup>29</sup>. They gave statistical reasoning for the cause of the association between PBA - CA and CA in linear regression. They also suggested removing the association through stage 2 correction of brain age delta, which was very similar to our RBA metric. Further, since decline in cognitive function is associated with brain aging<sup>30,31</sup>, we investigated the correlation between RBA and cognitive function as a proof that RBA was able to capture brain aging deviation from norm.

In this study, we aim to quantify how smoking, alcohol consumption, and genetic variants are associated with RBA. We analyzed brain-imaging data collected for 17,308 UK Biobank subjects who were cognitively normal and were of European ancestry. We first trained a model that produces RBA using data for 30% of the subjects. We then applied the trained model to the remaining 70% of the subjects (i.e., the evaluation set) and obtained RBA for those subjects. We further studied the association of RBA with smoking, alcohol consumption, and genetic variants, as well as interaction among those factors, using the evaluation set subjects.

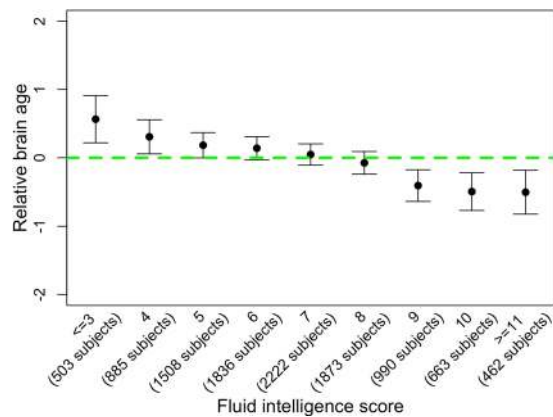
## Results

**Predicted brain age (PBA) and relative brain age (RBA).** We randomly split the data for 17,308 subjects with brain magnetic resonance imaging into training set ( $n = 5,193$ ) and evaluation set ( $n = 12,115$ ). Table 1 illustrates the demographic information for the subjects included in the training and evaluation sets. There was no significant difference in age, gender, smoking, and alcohol consumption (Supplementary Figs. 1, 2) between these two sets. We trained a model that produced the predicted brain age (PBA) and relative brain age (RBA) based on MRI measurements using training set subjects. We then applied this trained model to the evaluation set subjects (i.e., the evaluation set), and further obtained PBA and RBA for the evaluation set subjects (as illustrated in Fig. 1). The mean absolute error (MAE) between PBA and chronological age (CA) in the evaluation set was 3.8 years. The relationship between CA, PBA, and RBA for the evaluation set subjects is illustrated in Fig. 2 and Supplementary Figs. 3, 4. We carried out subsequent analyses using data of the evaluation set subjects.

**Cognitive function is negatively associated with RBA.** Subjects who performed better in the cognitive tasks had a lower RBA than that of those who performed worse. As shown in Fig. 3, Fluid intelligence score was



**Figure 2.** Relationship between chronological age and the predicted brain age. Subjects with positive relative brain age (RBA) are labeled with blue X's; subjects with negative RBA are labeled with red dots. The RBA of a specific subject is illustrated in pink color.



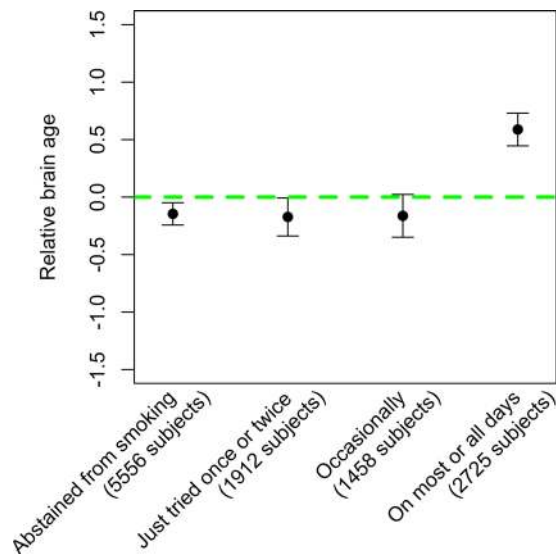
**Figure 3.** Relationship between Fluid intelligence score and relative brain age (RBA). Subjects with Fluid score of 0, 1, 2, and 3 ( $n = 28$ , 114, and 361, respectively) are grouped together. Subjects with Fluid score of 11, 12, and 13 ( $n = 334$ , 111, and 17, respectively) are grouped together, so that each group has more than 200 subjects.

negatively associated with RBA (Spearman's correlation =  $-0.07$ ,  $p$ -value =  $3E-13$ ; R-squared =  $0.005$ ). Further, a lower RBA was associated with a better performance in memorizing a specific command and in memorizing the position of matching card pairs, and a lower response time in identifying matching cards. Detailed results on are shown in Supplementary Figs. 5–9.

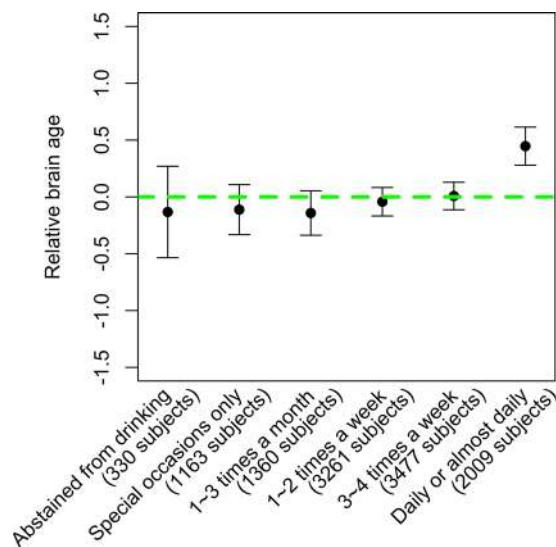
### Previous tobacco smoking and alcohol consumption are significantly associated with RBA.

Information of previous tobacco smoking frequency was collected for 11,651 of the evaluation set subjects during the visit for MRI scan. Regression analyses adjusting for gender and education showed that previous tobacco smoking frequency was statistically significantly associated with RBA (ANOVA F-test  $p$ -value  $< 2E-16$ , see Fig. 4). Pairwise comparisons showed that the most significant difference was between those who smoked on most or all days (with an average RBA of 0.6 years) and the rest of the smoking frequency categories (i.e., those who abstained from smoking, just tried once or twice, or occasionally), while there was no significant difference among the groups of subjects who didn't smoke on most or all days (Supplementary Table 1).

Information of current alcohol drinking frequency was collected for 11,600 of the evaluation set subjects during the visit for MRI scan. Regression analyses adjusting for gender and education showed that alcohol consumption frequency was statistically significantly associated with RBA (ANOVA F-test  $p$ -value =  $9E-6$ , see Fig. 5). Pairwise comparisons among groups with different alcohol consumption frequencies showed that the strongest difference was between the group who drank alcohol on most or all days (with an RBA of 0.4 years) and the rest of the alcohol drinking frequency categories (i.e., those who abstained from drinking, drank at special occasions



**Figure 4.** Relationship between previous tobacco smoking frequency and relative brain age.

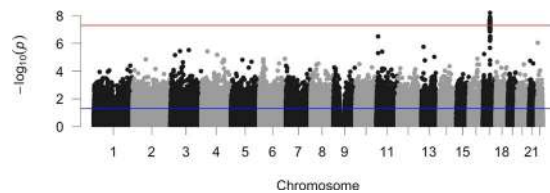


**Figure 5.** Relationship between alcohol intake frequency and relative brain age.

only, 1~3 times a month, 1~2 times a week, or 3~4 times a week), while the difference among groups who didn't drink on most or all days was insignificant (Supplementary Table 2).

Smoking and alcohol consumption amount were positively correlated and had an additive effect on RBA. Among the 2,327 subjects who smoked on most or all days and did not abstain from alcohol, the correlation between the two variables was 0.08 ( $p$ -value =  $9E-5$ ). We used a regression model with RBA as the response variable and with smoking amount, alcohol consumption amount, sex, and education as predictors. According to this model, each additional pack-year of smoking was associated with 0.03 years of increased RBA ( $p$ -value =  $2E-8$ ); each additional gram of alcohol consumption per day was associated with 0.02 years of increased RBA ( $p$ -value =  $6E-10$ ). The R-squared value of this model was 0.032. As a comparison, a model with only smoking amount as predictor and adjusted for sex and education had an R-squared of 0.018. A model with only alcohol consumption amount as predictor and adjusted for sex and education had an R-squared of 0.015. We also built a regression model with an interaction term between alcohol drinking and smoking. The interaction term was insignificant, indicating that there was insufficient evidence to support the presence of an interaction between alcohol drinking and smoking in affecting RBA.

**Association between single nucleotide polymorphisms and RBA.** We looked for single nucleotide polymorphisms (SNPs) that were associated with RBA within the evaluation set subjects. Multiple SNPs in a 2-Mb region on Chromosome 17 showed significant association with RBA (i.e.,  $p$ -value <  $5E-8$ ). Figure 6 is the Manhattan plot showing association  $p$ -value between SNPs and relative brain age across the genome.



**Figure 6.** Manhattan plot for the association p-values between genetic variants and relative brain age across the genome. The red line indicates the genome-wide significant threshold on p-value (i.e.,  $5E-8$ ). The blue line indicates p-value of 0.05.

Supplementary Table 3 lists the raw RBA-association p-values of 538,477 SNPs under analyses (doi: 10.5281/zenodo.3496206). A SNP located in MAPT gene showed the most significant association with RBA (SNP ID: Affx-13929237, p-value =  $6E-9$ ; see Supplementary Figs. 11, 12). We further built regression model to check if there was interaction between this SNP and smoking or alcohol consumption amount in association with RBA. The interaction term appeared to be insignificant in the model.

We also investigated the association between the dosage of APOE  $\epsilon 4$  risk allele, a major Alzheimer's disease risk factor, and RBA. We found that subjects with two copies of APOE  $\epsilon 4$  risk alleles had slightly higher RBA than subjects with zero or one copy of risk allele (Supplementary Fig. 13). However, the association between APOE risk allele dosage and RBA was only nominally significant (p-value = 0.03) and didn't reach significance threshold of genome-wide association studies.

## Discussion

Here we analyzed the brain imaging data collected for 17,308 UK Biobank subjects. We derived RBA metric using training set subjects, and further investigated the association of RBA with smoking, alcohol intake, and genetic variants using evaluation set subjects.

In our analyses, we first calculated PBA of a subject based on structural MRI data and then derived RBA, a metric that describes a subject's PBA relative to peers. RBA was calculated as the difference between PBA and EPBA (i.e.,  $RBA = PBA - EPBA$ ; see the methods section for details) of a person. As a comparison, in other studies where PBA was derived based on regression model, the difference between PBA and CA ( $PBA - CA$ , or BrainAGE) was used to indicate the brain aging status<sup>23,25,26</sup>. We observed that due to regression dilution, older subjects tend to have negative values of  $PBA - CA$ , while younger subjects tend to have positive values of  $PBA - CA$  (Fig. 1 and Supplementary Fig. 3). As a comparison, RBA was independent of CA. At all age ranges, roughly half of the subjects had positive RBA and half of the subjects had negative RBA (Fig. 1 and Supplementary Fig. 4). Therefore, the association analyses of RBA with lifestyle and genetic factors are not affected by chronological age.

We showed that subjects with positive RBA performed worse in various cognitive functions while subjects with negative RBA performed better. A relevant study reported that the biological brain aging accelerated in patients with cognitive impairment than in normal subjects<sup>27</sup>. Our findings further demonstrated that even among cognitively normal subjects, there was association between advanced brain age and declined cognitive function. We noticed that while the correlation between Fluid intelligence score and RBA was statistically significant, it was not strong. That is due to three main reasons. First, although aging is associated with declined cognitive function including Fluid intelligence, there are a lot of noises in this association (please also refer to our response to reviewer's 4th question). For example, researchers using UK Biobank have reported that there was a weak association between chronological age and Fluid intelligence score<sup>32</sup>. Second, RBA is independent of the chronological age. Therefore, although we expect that subjects with positive RBA to have worse cognitive function and subjects with negative RBA to have better cognitive function, we also acknowledge that subjects with the same RBA have a wide range of chronological age, causing large variation of Fluid intelligence score. Therefore, it is expected that RBA had a weak, yet significant, correlation with cognitive function. Third, subjects included in our analyses are cognitively normal. The association between RBA and cognitive function might be relatively weaker within these cognitively normal subjects as compared to a study in which subjects range from cognitively normal, mildly cognitive impaired, and severely cognitively impaired.

Our analyses of smoking and RBA indicated that subjects who had smoked on most or all days had a significantly higher RBA compared to subjects who smoked less often. That was consistent with previous studies, which showed significantly greater rate of atrophy in certain regions of the brains of smokers<sup>5-7</sup>. Our data also showed insignificant difference of RBA among the subjects who smoked occasionally, only tried once or twice, or abstained from smoking. This suggests that the detrimental effect of smoking on brain aging occurs mainly among those who smoked on most days.

Our analyses of alcohol intake frequency and RBA indicated that subjects who drank daily or almost daily had a significantly higher RBA compared to those who drank less frequently. Our finding was consistent with previous studies, which showed that heavy alcohol consumption was detrimental to the brain<sup>8,9,33</sup>. We did not find significant RBA difference among subjects who drank alcohol less frequently or those who abstain from drinking. It has been reported that a small dose of alcohol is associated with a reduced risk of cardiovascular disease, coronary heart disease and stroke<sup>34-36</sup>. Moreover, cardiovascular health and brain health are related. Researchers have found that cardiovascular risk factors like hypertension and heart disease are associated with increased brain white matter abnormalities and brain atrophy<sup>12,37</sup>. Therefore, a small amount of alcohol may have certain beneficial to brain health through contributing to the cardiovascular health. Gu *et al.*, have reported

that light-to-moderate total alcohol intake was associated with larger total brain volume in elderly subjects<sup>38</sup>. Nevertheless, our results didn't show RBA difference among subjects who drank alcohol less frequently or those who abstain from drinking. We also acknowledge that our observation would need to be further validated using an independent data set.

We also found genome-wide significant association between SNPs within a 2 Mb locus on chromosome 17 and RBA. The SNP showing the most significant association is located in MAPT gene. Previous studies also showed that mutations in MAPT, which encodes tau protein, are associated with dementia and Parkinson's disease<sup>39</sup>. Therefore, although the SNPs showing genome-wide significant association with RBA are located in a locus that covers multiple genes, it is likely that MAPT gene is a functional gene for brain aging. Further wet lab experiment focusing on MAPT gene and its regulatory region may be carried out to understand the etiology of brain aging. On the other hand, in previous studies, researchers have identified SNPs that showed genome-wide significant association with specific brain morphometrics. For example, SNP rs7294919 (candidate gene TESC) was associated with hippocampal volume; SNP rs945270 (candidate gene KTN1) was associated with putamen volume; SNP rs10784502 (candidate gene HMGA2) was associated with intracranial volume<sup>40,41</sup>. It is possible that since brain age was a summary statistic of the morphometrics of multiple brain regions, the associations between SNPs and specific brain regions did not get reflected. Although we only found SNPs on the 2 Mb chromosome 17 loci showing genome wide significant association with RBA, the SNP-level RBA association p-values can be used for future meta-analyses, where results from multiple genetic association studies are combined for identifying potentially more significant SNP-phenotype associations.

Several studies had been done previously to inspect the association between APOE  $\epsilon$ 4 risk allele, a major genetic risk factor for Alzheimer's disease (AD)<sup>42,43</sup>, and brain age. Cole *et al.*<sup>15</sup> looked at the association between APOE  $\epsilon$ 4 status and brain-predicted age difference (PAD) in 669 elderly subjects and reported no association between these two variables. Another study of 30 individuals with Down syndrome reported that APOE genotype did not significantly influence brain-PAD<sup>16</sup>. Lowe *et al.*<sup>17</sup> reported that APOE  $\epsilon$ 4 status did not have significant association with Brain Age Gap Estimation (BrainAGE) in healthy subjects, patients with AD or mild cognitive impairment. However, they did observe association between BrainAGE changing rates and APOE  $\epsilon$ 4 carrier status. In our analyses, we found that subjects with two copies of APOE risk alleles had slightly higher RBA than subjects with no risk allele or only one copy of risk allele, although the effect was not statistically significant. Therefore, the effect of APOE risk allele on brain aging is probably not strong within cognitively normal subjects.

Our study has some limitations. First, we used a linear regression model with LASSO to produce PBA based on structural MRI data. More sophisticated statistical approaches such as using principal component analyses for dimension reduction before LASSO regression, or using neural networks may help to improve the accuracy of PBA. Also, the combination of structural MRI and other types of brain imaging data (e.g., functional MRI, diffusion-weighted MRI) may help to improve the accuracy of PBA. A more accurate PBA would allow better estimation of RBA. Second, in our study, we investigated the association of brain age with tobacco smoking and alcohol consumption. Besides smoking and alcohol consumption, various environmental factors may be associated with brain age. For example, physical exercise and meditation had been reported to be associated with lower brain aging level<sup>44,45</sup>. Further, genetics also affects brain aging<sup>14</sup>. Therefore, the variation of RBA that can be explained by smoking and alcohol drinking amount was small (as reflected by the small R-squared in the regression model for quantifying the association of RBA with smoking and alcohol drinking amount). More studies can be done to help fully understand the factors associated with brain age. Third, we chose to use pack-years and grams of alcohol intake per day for assessing the smoking and drinking amount. There are alternative measurements for assessing smoking and drinking amount, which may yield slightly different findings<sup>46,47</sup>. Fourth, it is possible that more genetic variants that have strong effect on RBA do exist. However, these genetic variants may be missing from the current genotyping platform and are not detected through current analyses. Fifth, genetic predispositions are known to affect smoking and alcohol drinking behavior. For example, SNPs located in the region of alcohol-metabolizing enzyme genes are significantly associated with alcohol dependence<sup>48</sup>. A SNP located in the nicotinic receptor gene is significantly associated with number of cigarettes smoked per day<sup>49</sup>. Therefore, it is possible that genetic variants affect alcohol and nicotine consumption and indirectly affect the RBA. Sixth, a larger sample size would increase the power for identifying factors and interaction among them that are significantly associated with a specific trait. With increased number of UK Biobank subjects for whom brain-imaging data are available, a future study may reveal more about factors associated with brain aging.

In sum, we studied the association of brain age with smoking, alcohol consumption, and genetic variants using the data collected for 17,308 cognitively normal UK Biobank subjects. These results provided useful insights into how brain aging is associated with smoking and alcohol consumption. Our analyses only identified genetic variants within a 2 Mb locus on chromosome 17 to be significantly associated with brain aging. Further studies potentially with even larger sample sizes will be needed to provide a clearer picture of factors associated with brain aging.

## Materials and Methods

**Overview of UK biobank project.** The UK Biobank recruited ~500,000 subjects in the United Kingdom<sup>50</sup>. The participants have provided blood, urine and saliva samples. All participants have been genotyped. 20,000 participants scanned as of August 2018 were included in our study (including brain, heart, abdomen, bones and carotid artery). All participants had provided informed consent. The present analyses were conducted under data application number 25641.

**Magnetic resonance imaging (MRI) data.** Details of the structural brain MRI data, such as imaging hardware and acquisition protocols, are described elsewhere<sup>51,52</sup>. In our analyses, quality controlled structural MRI data was obtained for 21,345 subjects. We excluded 1,222 (5.7%) subjects with brain and nervous system

related illness, including cognitive impairment, neurological disorders or stroke, etc. (see Supplementary Table 4 for the list of diseases based on which subjects were excluded from our analyses). We further excluded 2,815 (13.2%) subjects with non-European ancestry (according to both self-reported ethnicity and principal component analyses on the genetic data). Brain imaging data of 17,308 subjects were used in our analyses. The age range of these participants is between 45.2 years and 80.7 years.

Brain morphometrics, including volume of cortical, subcortical and white matter regions, thickness and surface area of cortical regions, ventricle size, intracranial volume, etc., were obtained with FreeSurfer 6.0<sup>53</sup> based on the T1 MRI brain scans, with the Desikan-Killiany atlas. FreeSurfer is documented and freely available for download online (<http://surfer.nmr.mgh.harvard.edu/>). Supplementary Table 5 lists the 403 brain morphometric measurements used in our analyses.

**Cognitive function.** We used the data of cognitive function in its original form, which was collected during the visit for MRI scan. All subjects performed specific tasks as instructed by a computer. To be specific, the Fluid intelligence score indicates the capacity to solve problems that require logic and reasoning ability. It was based on subjects' performance in identifying the largest number, calculating family relationship, interpolating word, etc. For the prospective memory task, subjects were asked to memorize a command in the middle of the cognitive tests and perform it at the end of the test. In the reaction time test, subjects were asked to press a snap-button when two cards displayed on the computer screen matched. Mean time to correctly identify matches was recorded. In the pairs matching test, subjects were asked to memorize the position of matching pairs of cards. The number of correct pairs identified was recorded. More details of the tasks for assessing cognitive function can be found on the UK Biobank website (<http://www.ukbiobank.ac.uk/>).

**Education.** We used the information of education qualification collected during the visit for MRI scan. The qualification variable has multiple categories based on a British system. We collapsed it into two categories indicating whether or not a subject held a college or university degree, as used in the paper by Cox *et al.*<sup>54</sup>. There was a significant association between education and RBA ( $p$ -value = 0.009, Supplementary Fig. 10). Therefore, we also adjusted for education when assessing the association of RBA with smoking, alcohol consumption, and genetic variants.

**Tobacco smoking history and alcohol intake.** We used the information of smoking history and alcohol intake status that was collected during the visit for MRI scan. The smoking and alcohol intake frequency categories used in our analyses were as reported in the UK Biobank questionnaire. The smoking pack-years was defined as the number of cigarettes smoked per day/20 multiplied by the number of years of smoking. The alcohol intake amount was calculated as described in the paper by Piumatti *et al.*<sup>56</sup>. Alcohol consumption per day for a specific type of drink was calculated as the number of drinks consumed per day multiplied by the number of grams of alcohol contained in one drink. The total amount of alcohol consumption per day was the summation of the alcohol amount from all types of drinks. More details can be found on the UK Biobank website (<http://www.ukbiobank.ac.uk/>).

**Genotype data.** Details of the genotyping and genotype calling procedures are described elsewhere<sup>55</sup>. Quality-controlled genotype data was obtained for 538,477 autosomal SNPs genotyped for 11,900 evaluation set subjects. Our quality control on SNPs ensured that all SNPs had missing rate less than 0.02 and passed Hardy-Weinberg exact test (i.e., Hardy-Weinberg equilibrium  $p$ -value  $\geq 1E-6$ ). Quality control on the samples ensured that all subjects had genotyping rate greater than 0.98 and had heterozygosity rate within  $\pm 3$  standard deviation, had matched reported gender and genetic gender, and were of European ancestry (according to both self-reported ethnicity and genetic ethnicity based on principal component analyses). Related individuals (i.e., kinship coefficient  $>0.1$ ) were further removed.

**Obtaining predicted brain age (PBA) and relative brain age (RBA) based on structural MRI data.** Predicted brain age (PBA) is a metric describing how old a person's brain appears based on a brain scan at a single time-point. Relative brain age (RBA) is a metric indicating if a person's brain has experienced accelerated or decelerated aging compared to peers. It captures the deviation of a person's brain structural aging from the population's normal pace.

We trained a model for obtaining PBA and RBA based on MRI data using training set subjects. To be specific, we randomly split the brain imaging data of 17,308 subjects into training and evaluation sets. Our rationale for picking 30% (5,193) of the subjects as the training set and the remaining 70% (12,115) as the evaluation set was to balance the need for accurately training a model to predict brain age and the need for a large number of subjects in the evaluation set for evaluating the association of RBA and the factors of interest.

The model for obtaining PBA and RBA is trained as follows. We first trained a model obtaining predicted brain age (PBA) based on MRI data using data of the training set subjects. To be specific, we built a linear regression model with Lasso regularization for predicting brain age using R package *glmnet*<sup>56,57</sup>. In the model, the chronological age was the response variable, and 403 brain quantitative measures derived using *FreeSurfer* were used as predictors. During model training, the Lasso parameter, lambda, was selected based on an internal cross validation using *glmnet*. We did not do any pre-selection on the predictors, since the training set sample size was sufficiently large relative to the number of predictors in the model. The mean absolute error (MAE) between PBA and chronological age in the training set was 3.5 years. We observed that due to regression dilution<sup>28</sup>, the difference between PBA and CA (i.e., PBA - CA) was negatively associated with CA. The older subjects tended to have negative PBA - CA, while the younger subjects tended to have positive PBA - CA (See Fig. 2 and Supplementary Fig. 3). Therefore, after obtaining PBA for each subject, we further calculated RBA. RBA is

defined as the difference between PBA and expected PBA given a subject's chronological age (i.e.,  $RBA = PBA - \text{Expected}(PBA|CA)$ ). Here,  $\text{Expected}(PBA|CA)$ , or EPBA, was obtained through building a regression model where CA was the predictor and PBA was the response variable. In that way, RBA is independent of CA. At each age range, there were roughly half of the subjects with positive RBA and half of the subjects with negative RBA (Fig. 1, Supplementary Fig. 4). A subject with positive RBA has a brain that appears older than those of peers, while a subject with negative RBA has a brain that appears younger. Since we linear operations were used to derive RBA based on PBA and CA, the unit of RBA is year.

After training the model for obtaining PBA and RBA using the training set data, we applied it to the evaluation set and carried on association analyses.

### Quantifying the association of RBA with previous tobacco smoking amount and alcohol intake amount.

We quantified the association between previous tobacco smoking amount, alcohol intake amount, and RBA using a two-step regression model adjusting for gender and education. We first built a linear regression model using data of 2,327 evaluation set subjects who previous smoked daily or almost daily and did not abstain from drinking alcohol. We then identified subjects with large Cook's distance as potential influential observations (i.e., subjects with Cook's distance greater than 3\* the mean Cook's distance of all the subjects). We excluded these influential observations, fitted a second linear regression model, and reported results based on the second regression model. In total, data of 2,174 non-influential observations were used in the second-step regression.

**Testing the association between genetic variants and RBA.** We used PLINK<sup>58</sup> linear regression model for genotypic test, adjusting for gender, education, and first five genetic principal components of ancestry, to test the association between SNPs and RBA.

Received: 18 April 2019; Accepted: 6 December 2019;

Published online: 30 January 2020

### References

1. Ortman, J., Velkoff, V. & Hogan, H. An Aging Nation: The Older Population in the United States (2014).
2. Lindenberger, U. Human cognitive aging: corrigir la fortune? *Science* **346**, 572–8 (2014).
3. Jack, C. R. Jr. *et al.* Age, Sex, and APOE epsilon4 Effects on Memory, Brain Structure, and beta-Amyloid Across the Adult Life Span. *JAMA Neurol* **72**, 511–9 (2015).
4. Andersen, K. *et al.* Gender differences in the incidence of AD and vascular dementia: The EURODEM Studies. EURODEM Incidence Research Group. *Neurology* **53**, 1992–7 (1999).
5. Durazzo, T. C., Insel, P. S. & Weiner, M. W. & Alzheimer Disease Neuroimaging, I. Greater regional brain atrophy rate in healthy elderly subjects with a history of cigarette smoking. *Alzheimers Dement* **8**, 513–9 (2012).
6. Duriez, Q., Crivello, F. & Mazoyer, B. Sex-related and tissue-specific effects of tobacco smoking on brain atrophy: assessment in a large longitudinal cohort of healthy elderly. *Front Aging Neurosci* **6**, 299 (2014).
7. Gallinat, J. *et al.* Smoking and structural brain deficits: a volumetric MR investigation. *Eur J Neurosci* **24**, 1744–50 (2006).
8. Pfefferbaum, A. *et al.* Brain gray and white matter volume loss accelerates with aging in chronic alcoholics: a quantitative MRI study. *Alcohol Clin Exp Res* **16**, 1078–89 (1992).
9. Asensio, S. *et al.* Magnetic resonance imaging structural alterations in brain of alcohol abusers and its association with impulsivity. *Addict Biol* **21**, 962–71 (2016).
10. Ettinger, U. *et al.* Effects of acute nicotine on brain function in healthy smokers and non-smokers: estimation of inter-individual response heterogeneity. *Neuroimage* **45**, 549–61 (2009).
11. Gold, M., Newhouse, P. A., Howard, D. & Kryscio, R. J. Nicotine treatment of mild cognitive impairment: a 6-month double-blind pilot clinical trial. *Neurology* **78**, 1895; author reply 1895 (2012).
12. Almeida, O. P. *et al.* Coronary heart disease is associated with regional grey matter volume loss: implications for cognitive function and behaviour. *Intern Med J* **38**, 599–606 (2008).
13. Gianaros, P. J., Greer, P. J., Ryan, C. M. & Jennings, J. R. Higher blood pressure predicts lower regional grey matter volume: Consequences on short-term information processing. *Neuroimage* **31**, 754–65 (2006).
14. Cole, J. H. *et al.* Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *Neuroimage* **163**, 115–124 (2017).
15. Cole, J. H. *et al.* Brain age predicts mortality. *Mol Psychiatry* **23**, 1385–1392 (2018).
16. Cole, J. H. *et al.* Brain-predicted age in Down syndrome is associated with beta amyloid deposition and cognitive decline. *Neurobiol Aging* **56**, 41–49 (2017).
17. Lowe, L. C., Gaser, C. & Franke, K. & Alzheimer's Disease Neuroimaging, I. The Effect of the APOE Genotype on Individual BrainAGE in Normal Aging, Mild Cognitive Impairment, and Alzheimer's Disease. *PLoS One* **11**, e0157514 (2016).
18. Dick, D. M. Gene-environment interaction in psychological traits and disorders. *Annu Rev Clin Psychol* **7**, 383–409 (2011).
19. Hollman, A. L., Tchounwou, P. B. & Huang, H. C. The Association between Gene-Environment Interactions and Diseases Involving the Human GST Superfamily with SNP Variants. *Int J Environ Res Public Health* **13**, 379 (2016).
20. Hunter, D. J. Gene-environment interactions in human diseases. *Nat Rev Genet* **6**, 287–98 (2005).
21. Peila, R. *et al.* Joint effect of the APOE gene and midlife systolic blood pressure on late-life cognitive impairment: the Honolulu-Asia aging study. *Stroke* **32**, 2882–9 (2001).
22. Anisimov, V. N. *et al.* Metformin slows down aging and extends life span of female SHR mice. *Cell Cycle* **7**, 2769–73 (2008).
23. Franke, K., Ziegler, G., Kloppel, S., Gaser, C. & Alzheimer's Disease Neuroimaging, I. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. *Neuroimage* **50**, 883–92 (2010).
24. Cole, J. H. & Franke, K. Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends Neurosci* **40**, 681–690 (2017).
25. Franke, K., Gaser, C., Manor, B. & Novak, V. Advanced BrainAGE in older adults with type 2 diabetes mellitus. *Front Aging Neurosci* **5**, 90 (2013).
26. Nenadic, I., Dietzek, M., Langbein, K., Sauer, H. & Gaser, C. BrainAGE score indicates accelerated brain aging in schizophrenia, but not bipolar disorder. *Psychiatry Res* **266**, 86–89 (2017).
27. Liem, F. *et al.* Predicting brain-age from multimodal imaging data captures cognitive impairment. *Neuroimage* **148**, 179–188 (2017).
28. Hutcheon, J. A., Chioloro, A. & Hanley, J. A. Random measurement error and regression dilution bias. *BMJ* **340**, c2289 (2010).
29. Smith, S. M., Vidaurrre, D., Alfaro-Almagro, F., Nichols, T. E. & Miller, K. L. Estimation of brain age delta from brain imaging. *Neuroimage* **200**, 528–539 (2019).
30. Creasey, H. & Rapoport, S. I. The aging human brain. *Ann Neurol* **17**, 2–10 (1985).



31. Luo, L. & Craik, F. I. Aging and memory: a cognitive approach. *Can J Psychiatry* **53**, 346–53 (2008).
32. Klinedinst, B. S. *et al.* Aging-related changes in fluid intelligence, muscle and adipose mass, and sex-specific immunologic mediation: A longitudinal UK Biobank study. *Brain Behav Immun* **82**, 396–405 (2019).
33. Shokri-Kojori, E., Tomasi, D., Wiers, C. E., Wang, G. J. & Volkow, N. D. Alcohol affects brain functional connectivity and its coupling with behavior: greater effects in male heavy drinkers. *Mol Psychiatry* **22**, 1185–1195 (2017).
34. Corrao, G., Rubbiati, L., Bagnardi, V., Zambon, A. & Poikolainen, K. Alcohol and coronary heart disease: a meta-analysis. *Addiction* **95**, 1505–23 (2000).
35. Ronksley, P. E., Brien, S. E., Turner, B. J., Mukamal, K. J. & Ghali, W. A. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *BMJ* **342**, d671 (2011).
36. Piumatti, G., Moore, S., Berridge, D., Sarkar, C. & Gallacher, J. The relationship between alcohol use and long-term cognitive decline in middle and late life: a longitudinal analysis using UK Biobank. *J Public Health (Oxf)* (2018).
37. Kappus, N. *et al.* Cardiovascular risk factors are associated with increased lesion burden and brain atrophy in multiple sclerosis. *J Neurol Neurosurg Psychiatry* **87**, 181–7 (2016).
38. Gu, Y. *et al.* Alcohol intake and brain structure in a multiethnic elderly cohort. *Clin Nutr* **33**, 662–7 (2014).
39. Goedert, M. NEURODEGENERATION. Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled Abeta, tau, and alpha-synuclein. *Science* **349**, 1255555 (2015).
40. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224–9 (2015).
41. Stein, J. L. *et al.* Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* **44**, 552–61 (2012).
42. Saunders, A. M. *et al.* Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* **43**, 1467–72 (1993).
43. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452–8 (2013).
44. Luders, E., Cherbuin, N. & Gaser, C. Estimating brain age using high-resolution pattern recognition: Younger brains in long-term meditation practitioners. *Neuroimage* **134**, 508–513 (2016).
45. Steffener, J. *et al.* Differences between chronological and brain age are related to education and self-reported physical activity. *Neurobiol Aging* **40**, 138–44 (2016).
46. Neuner, B. *et al.* Modeling smoking history: a comparison of different approaches in the MARS study on age-related maculopathy. *Ann Epidemiol* **17**, 615–21 (2007).
47. Wood, M. A., Kaptoge, S. & Butterworth, S. A. Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599912 current drinkers in 83 prospective studies. *The Lancet* **391** (2018).
48. Hart, A. B. & Kranzler, H. R. Alcohol Dependence Genetics: Lessons Learned From Genome-Wide Association Studies (GWAS) and Post-GWAS Analyses. *Alcohol Clin Exp Res* **39**, 1312–27 (2015).
49. The Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441–7 (2010).
50. Allen, N. E., Sudlow, C., Peakman, T., Collins, R. & Biobank, U. K. UK biobank data: come and get it. *Sci Transl Med* **6**, 224ed4 (2014).
51. Miller, K. L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* **19**, 1523–1536 (2016).
52. Smith, S., Alfaro-Almagro, F. & Miller, K. UK Biobank Brain Imaging Documentation (2017).
53. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–81 (2012).
54. Cox, S. R. *et al.* Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nat Commun* **7**, 13629 (2016).
55. UKBiobank. Genotyping and quality control of UK Biobank, a large-scale, extensively phenotyped prospective resource (2015).
56. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1–22 (2010).
57. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing (2012).
58. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–75 (2007).

## Acknowledgements

We thank Dr. Bo Chen for helpful discussions on the data analyses procedure. We also acknowledge the contributions of members of the UK Biobank coordinating center. This work was supported by grants P41EB015922, U54EB020406 and R01MH094343 of the National Institutes of Health.

## Author contributions

K.N. wrote the main manuscript text and prepared figures and tables. L.Z. acquired and arranged data from UK Biobank. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-56089-4>.

**Correspondence** and requests for materials should be addressed to A.W.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020